The Evolution of Data Publication and the Role of Persistent Identifiers and Linked Open Data in Dynamic Data Mobilization

Peter L. Pulsifer[1*]

[1*] *National Snow and Ice Data Center, University of Colorado, 449 UCB, Boulder, Colorado, 80309, USA*
Email: pulsifer@nsidc.org

Summary. Data publication and citation models have existed for millennia. In recent decades, Information and Communications Technologies have transformed the way that we publish and cite data. A Linked Open Data approach combined with a variety for different persistent identifier schemes and accessed through a semantic web model is becoming a powerful system for mobilizing data. This type of approach will be necessary as data systems evolve and must be considered in our discussions about sharing data across disciplines.

Keywords. Linked Open Data, Persistent Identifiers, Semantic Web, Interoperability, History of Science.

1. Introduction

Since before the advent of the printing press in the fifteenth century, humans have collected, organized and disseminated knowledge to better understand their environment and society. For millennia, Indigenous Peoples have synthesized empirical observations and shared knowledge orally, and more recently using writing and digital technologies [1]. Before Current Era, the Library at Alexandria held and managed thousands of documents and it loss is recognized as an example of the loss of an era of knowledge. The International Polar Year (1882-83) marks a milestone in formal data management and publication for the Polar Regions, as was the formation of the World Data System emerging from the International Geophysical Year (1957-58). These systems have served us well, however in the era of digital information and communications technologies, there is a movement towards a new form of data publication based on identifiers such as the Digital Object Identifier (DOI) [2]. The appropriateness of "publication" as a conceptualization for referencing and accessing published knowledge has been debated [3]. In this paper we examine established and emerging mechanisms for publishing and citing data and argue that "Linked Open Data" (LOD) models will continue to evolve as a foundational layer in the way that humans manage knowledge.

2. Linked Open Data

Linked Open Data is a term used to describe exposing, sharing, and connecting information resources on the Web using persistent identifiers (PIDs) and the Semantic Web model [4]. DOIs are well known PIDs, however there are many others being used to positively identify data resources [5]. The LOD and Semantic Web model provide a framework for documenting, understanding and combining information resources with PIDs that goes well beyond simple linking.

3. Linked Open Data, Data Publication and Citation

In contemporary research and society, data are often dynamic as are the information products generated from the transformation and mediation of data resources. An effective data publication and citation model will need to consider how to deal with dynamic data (including metadata) that are represented and identified in different ways. Relying solely on manual, deliberate, human-authored data publications will result in a significant paucity in

available information. This is particularly important as we see the emergence of sensor webs and The Internet of Things as sources of observational data.

4. Conclusions

Ultimately, our data publication and citation model will be based on LOD based on PIDs, the concepts of the Semantic Web (however they evolve), and ultimately techniques such as machine learning. This will allow for the mobilization of dynamic data through various forms of mediation, ultimately serving a myriad of applications of value to society. Failure to consider these emerging trends now may limit the success of the movement towards data publication and citation as it is often defined.

Acknowledgments. The author acknowledges the support of the U.S. National Science Foundation.

References

1. Pulsifer, P., Gearheard, S., Huntington, H. P., Parsons, M. A., McNeave, C., McCann, H. S.. The role of data management in engaging communities in Arctic research: overview of the Exchange for Local Observations and Knowledge of the Arctic (ELOKA). *Polar Geogr.,* 35, 271–290, 2012
2. Lawrence, B., Jones, C., Matthews, B., Pepler, S., Callaghan S., Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation,* 6(2), 2011
3. Parsons, M. A., Fox, P. A., "Is data publication the right metaphor?", *Data Science Journal* , 12, 2013.
4. Pulsifer, P. L., Brauen, G., Geo-semantic Web. Kitchin, R., Lauriault, T. P. Wilson, M. W. (*Eds.),* in *Understanding Spatial Media*. Sage, London, isbn: 9781473949683, 136-148, 2017
5. Duerr, Ruth, E., et al., "On the utility of identification schemes for digital earth science data: an assessment and recommendations." *Earth Science Informatics*, 4.3, 2011, 139, 2011